DOCUMENT RESUME

ED 087 806 TM 003 429

AUTHOR Powers, James E.
TITLE The Application of Bayes' Formula to Expectancy Tables.
PUB DATE Oct 73
NOTE 17p.; Paper presented at Northeastern Educational Research Association Meeting (Ellenville, New York, October, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Bayesian Statistics; *College Bound Students; *Expectancy Tables; *Guidance; Individual Characteristics; Prediction; Predictor Variables

ABSTRACT

Recently the application of Bayesian statistical procedures within the field of guidance and counseling has been receiving attention. This paper justifies and illustrates a simple, direct application of Bayes' Formula to expectancy tables. To make the application a counselor must specify prior probabilities that a student will attain success (defined appropriately). He may then take the probabilities within the expectancy table as likelihoods, apply Bayes' formula and obtain a posterior distribution of success for that student. This procedure allows a counselor the advantage of bringing to bear his personal knowledge of those students with whom he works. It will enable him to consider such diverse things as whether a student is an over-achiever, has just recovered from an extended illness or even the fact that there was a death in the student's family recently. Hopefully, the method will allow counselors to better aid their clients. A complete example demonstrating the application is included. (Author)

ED 087806

TM 003 429

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPPESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

The Application of Bayes' Formula to Expectancy Tables

James E. Powers

State University of New York at Albany

Presented at the 1973 meeting of the

Northeastern Educational Research Association

Fallsview Hotel

Ellenville, New York

ED 087806

TM 003 429

Abstract

Recently the application of Bayesian statistical procedures within the field of guidance and counseling has been receiving attention. This paper justifies and illustrates a simple, direct application of Bayes' Formula to expectancy tables.

To make the application a counselor must specify prior probabilities that a student will attain success (defined appropriately). He may then take the probabilities within the expectancy table as likelihoods, apply Bayes' formula and obtain a posterior distribution of success for that student.

This procedure allows a counselor the advantage of bringing to bear his personal knowledge of those students with whom he works. It will enable him to consider such diverse things as whether a student is an over-achiever, has just recovered from an extended illness or even the fact that there was a death in the student's family recently. Hopefully, the method will allow counselors to better aid their clients.

A complete example demonstrating the application is included.

Introduction

Recently the application of Bayesian statistical procedures within the field of guidance and counseling has been receiving attention. A series of three papers by Lindley (1969a, 1969b, 1970) focuses on the problem of colleges' selection of students from various high schools when the purpose is to choose those who will have a high grade point average at the end of their college careers. He employs a regression model using normal distributions. Some similar work by Owen (1969) employs a Poisson model.

Novick, Jackson, Thayer and Cole (1971) applied the method suggested by Lindley to data from the Basic Research Service of the American College Testing Program, finding that the Bayesian procedure led to a notable increase in predictive effeciency, at least in a group of homogeneous colleges when only small samples of students were available. In particular they found that in these situations errors of prediction were the same for a 25% sample of students when the Bayesian method was employed as they were for a 100% sample using the traditional least squares method.

In a paper discussing the use of educational tests in guidance services, Novick and Jackson (1969) suggest that Bayesian procedures may increase the effectiveness of counseling services since they show "an increased sensitivity when data are scarce and a resulting ability to discard obsolete data and thus keep up with current trends (p. 47)." These same authors also point outtthat this is essentially a clerical task that could be done centrally, thus freeing counselors to spend more time helping students individually (p. 47).

The prediction procedures suggested by Lindley are aimed at the problem of colleges' selection of students. A problem of equal or perhaps greater importance is that of the high school student who must decide which of several colleges to attend. Since many students are acdepted by more than one college, the decision is very real and perplexing. To arrive at the best course of action these students require easily understood and meaningful data.

Novick and Jackson (1969) suggest that a student be given estimates of his future grade point average for both his first year in college and for his entire program. They also feel that he should be given estimates of the probabilities that he will complete both his first year and entire program. The latter information is not generally disclosed in current practice (p. 15).

These same authors also point out one of the commonly mentioned objections to the use of statistical prediction procedures--that they are not personal, but are usually applied to groups. They write:

> "Formal classification models reflect the point view that if assignments are good, on the average, then a satisfactory state of affairs has been attained. The student, however, is unconcerned with such average good. If he perceives that he belongs to some subgroup for which, on the average, poor assignment decisions are made, it will not comfort him to know that the system works well for almost everybody else (p. 5)."

A way to provide students with meaningful, understandable estimates of the previously mentioned quantities, which will also be personal and easily obtained by gkuidance counselors, is available through the application of Bayes' formula to expectancy tables.

Bayes' Formula

In order to employ Bayes' formula it is necessary that one specify prior probabilites that the events of concern will occur. These prior probabilites should represent everything that is known up to that point in time.

Within the context of expectancy tables this means that a counselor must specify a priori the probability that a particular student will succeed. To do this, he must bring to bear all of the information available about the student and quantify it to arrive at the required prior probability. The only restriction upon the information considered is that it be relevant to whether or not the student will succeed. Obviously such things as scores on standardized tests and high school grades are relevant. However, a counselor is not confined to these usual predictors. He may also consider students' motivation, study habits, desire to attend a particular college and other facts that are usually not included if forming expectancy tables.

Thus it it the prior probability that permits the information contained in the table to be brought to a personal level for a given student since it allows for the consideration of data peculiar to that student.

Having specified the prior probability, it is then necessary to obtain the second quantity needed to employ the formula, called the likelihoods. Generally the likelihoods are conditional probabilities that the events of concern will happen given some set of observed data. When dealing with expectancy tables, the likelihood is the probability that the student will succeed given some predictive measure and is furnished by the entries in the body of the table.

These two quantitues (prior probability and likelihood) are then combined through Bayes' formula to obtain the posterior probabilities. The posterior probabilities now represent the probability that the events of concern will occur at this new point in time. They incorporate all available information (prior information plus information from data). With expectancy tables, the posterior probability is the probability that the student will succeed based upon all data relevant for him.

Therefore, the resulting probability of success furnished to the student is not simply an impersonal number based upon a large group of students who obtained the same score on that particular predictive measure, but a personal probability that incorporates many of his individual characteristics as well. This estimate of his personal probability should be useful and meaningful and should aid him in making a deciison about which college to attend.

For these purposes this posterior probability of success may be calculated through the application of Bayes' formula written as

$$(1) \quad P(S) = \frac{xy}{xy + (1-x)(1-y)}$$

where P(S) represents the posterior probability of success for the student, x represents the prior probability for that student and y is the likelihood obtained from the expectancy table. The derivation of this specialized formula is given in the Appendix.

The application of the formula is not mathematically difficult; however, it will no doubt require more time and effort (and perhaps place a greater responsibility on the counselor) than the usual use of expectancy tables, mostly due to the need to acquire and quantify prior

information. Of course, most counselors should have considerable information readily available regarding their students, and, as shall be shown in a later section of this paper, great accuracy in the prior probabilities is not always necessary nor will the use of the formula be necessary for all students.

Hopefully, the gain in usefulness to the student of the data obtained in this fashion will far outweigh the slight additional labor necessary.

An Example

Suppose that a counselor is working with students and employing the expectancy table shown in Table 1.

First consider the hypothetical example of John. Before talking with John, the counselor must quantify the prior information he has available for him. Quite probably this will entail reviewing John's high school grades and scores on previously administered standardized aptitude and achievement tests. However, other more personal information about John may be available, theough the counselor's own previous contact with him or from his teachers, and should also be considered.

To quantify the available prior information the counselor should probably first decide whether John's probability of success is above or below .5 and then decide how much above or below.

In this case let us first suppose that John has consistently been an honor student throughout his high school career and has consistently scored above the 90th percentile on aptitude and achievement tests. Further, John comes from a family of comfortable income and has been awarded a scholarship, so his finances are insured. In short, John

Table 1

Percentile Rank on a Scholastic Aptitude Test and College Semesters Completed[1]

Percentile Rank	Semester 1	2	3	4	5	6	7	8
90-	100	98	97	95	93	90	88	86
80-89	100	98	96	93	92	91	86	83
70-79	100	97	94	91	90	89	84	80
60-69	100	96	92	88	86	84	81	77
50-59	99	95	91	87	85	82	79	74
40-49	98	91	90	86	83	78	75	68
30-39	97	88	86	82	79	72	67	62
20-29	95	81	80	77	74	66	59	56
10-19	92	75	73	67	64	58	55	51
0-9	88	68	64	60	57	54	51	47

[1] The numbers in the table represent the percentage of students in the specific range of percentile ranks who completed the indicated number of semesters.

appears certain to be a collegiate success. The counselor might then assign a prior probability of .90 that John will complete eight semesters of college. He does not assign a 1, indicating certainty, since he is aware that unforeseen events might occur. John might get married and leave school, a family or personal tragedy might force him to leave, or some other event might prevent him from completing his college program.

Now suppose that on the predictive measure employed in the expectancy table John scores at the 95th percentile. We have the prior probability of success, x, as .90 and the likelihood of success, y, from the table as .86. Employing Bayes' formula now gives the posterior probability of success, P(S), as

$$P(S) = \frac{xy}{xy + (1-x)(1-y)} = \frac{.9(.86)}{.9(.86) + .1(.14)} = .98$$

Thus it appears that John is almost certain of success when all available information is utilized, whereas, had only the data from the table been considered, his probability of success would have been reported to him as only .86.

So it may be reported to John that the probability that he personally will complete his college program is .98.

In reporting such data to students it may be more meaningful to them in terms of odds. Here the odds are 98 to 2 or 49 to 1 that John will be successful. (Note that using solely the table would give odds of only 86 to 14 or 6.1 to 1.)

Suppose, instead, that the prior information about John were as follows. Although his scores on aptitude tests have been consistently high, his high school grades have been only marginally passing. Upon

speaking with several teachers, the counselor finds that John cuts classes a lot and simply does not do the required work. However, John does want to go to college.

In this case the counselor might evaluate John's prior probability of completing eight semesters of college at only .25 if he continues to neglect his coursework. However, if John were to exert some effort, the prior probability might be set at .75.

Assuming that John scores at the 81th percentile of the predictive instrument used in constructing the table and employing Bayes' formula with each of the prior probabilities, we have

$$P(S) = \frac{.25(.83)}{.25(.83) + .75(.17)} = .62$$

and $$P(S) = \frac{.75(.83)}{.75(.83) + .25(.17)} = .94$$

or in terms of odds, 1.6 to 1 and 15.7 to 1 in favor of him completing his college program. John may now be advised that with his present study habits he has only a slightly better than even chance of getting through college. However, if he were to apply himself, he has an extremely good chance.

In the event that the counselor had absolutely no information about John, the only available prior estimate of success would be .5 since he would have no evidence one way or the other. In this case the posterior estimate is simply the likelihood from the table due to the cancellation of the prior probabilities in the formula, i.e.,

$$P(S) = \frac{.5y}{.5y + .5(1-y)} = \frac{y}{y + 1-y} = y$$

However, it does not seem very likely that a counselor would ever have absolutely no prior information about a student.

Discussion

As suggested earlier in this paper, it does not seem that extreme accuracy is needed in the specification of the prior probability of success for this application of Bayes' formula.

For example, suppose a student scores above the 90th percentile on the predictive measure. This yields a likelihood of .86 (from Table 1) that he will complete eight semesters of college. If the prior probability of success for this student is assessed at .9, then the posterior probability is .98. If the prior probability were set at .8, the posterior would be .96 and if the prior were .7, the posterior would be .93.

It does not seem likely that a student's decision would be altered depending on whether his probability of completing the program was reported to him as .93, .96 or .98. In fact, quite conceivably a student would not be differentially influenced by probabilities as far as .10 apart. Would a student select College A over College B because his probability of success at A was .85 while at B it was only .75? It does not seem so.

With this point in mind, Table 2 was prepared. This table shows the posterior probabilities of a student's completing eight semesters of college for various combinations of prior probabilities and likelihoods. The posterior probabilities of completing eight semesters were selected for display because it seems that these will be the ones of most concern to students. It is, of course, possible to prepare similar tables for any number of semester..

To use Table 2, a counselor need only (1) specify the prior for eight semesters; (2) read the likelihood from Table 1 and round it to the nearest tenth; and (3) then read the posterior for that student from Table 2.

Table 2

Posterior Probabilities of Completing Eight Semesters of College for Selected Prior Probabilities and Likelihoods

Prior Probability	Likelihood								
	9	8	7	6	5	4	3	2	1
9	99	97	95	93	90	86	70	69	50
8	97	94	90	86	80	73	63	50	31
7	95	90	84	78	70	61	50	37	21
6	93	86	78	69	60	50	39	27	14
5	90	80	70	60	50	40	30	20	10
4	86	73	61	50	40	31	22	14	07
3	79	63	50	39	30	22	16	10	05
2	69	50	37	27	20	14	10	06	03
1	50	31	21	14	10	07	05	03	01

For example, suppose the prior of completing eight semesters for a particular student is specified as .8. Assuming the student scores at the 47th percentile on the predictive measure, a likelihood of .68 is obtained from Table 1. Entering Table 2 with a prior of .8 and a likelihood of .7 now gives the posterior of eight semesters for that student as .90.

This posterior is abviously not completely accurate due to the rounding of the likelihood. However, it should be sufficiently precise for this purpose. Calculated by formula this posterior is .89, a difference hardly worth noting.

In the event that a counselor is not confident about his evaluation of the prior, Table 2 makes it quick and easy for him to see that if the prior had been taken as .9 or .7, the resulting posteriors would have been .95 and .84 respectively. Thus he can immediately see that the magnitude of the differences is not substantial.

Tables such as this should greatly facilitate the use of this method due to the savings in time and effort afforded by the elimination of the arithmetic computations.

Summary

It is suggested that the use of Bayes' formula in conjunction with expectancy tables will provide personal data to prospective college students. Hopefully this data will be a greater aid in college selection than is data usually applied to groups.

Bayes' formula is discussed briefly and the use of a simplified version of the formula with expectancy tables is illustrated with a hypothetical example.

Lastly a table of posterior probabilities is presented and discussed. Tables such as this enable counselors to simply read the posterior probability that a student will succeed once they have evaluated the prior probability of success and obtained the likelihood of success from an expectancy table.

In conclusion an expectancy table is one counseling device to which Bayesian methodology can be directly and simply applied. Tables of posterior probabilities eliminate the need to evaluate even the simplified version of Bayes' formula in most cases and should greatly facilitate the use tf the method. Almost certainly there are many other counseling instruments to which Bayesian methods can be easily applied to furnish clients with valuable, personal information.

Appendix

Bayes' formula, in the discrete case, is given by

$$(1)\quad P(E_i|E) = \frac{P(E_i)P(E|E_i)}{\sum_{i=1}^{k} P(E_i)P(E|E_i)}$$

where

$P(E_i)$ is the prior probability that the ith event will occur.

$P(E|E_i)$ is the probability (likelihood) that the event E will occur, given that the ith event occurs.

$P(E_i|E)$ is the posterior probability that the ith event will occur, given that the event E has occured.

k is the number of events possible.

Within the application of this formula to an expectancy table, there are only two events of concern: E_1 is the event that the student will succeed; and E_2 is the event that he or she will not. Since the two events are mutually exclusive and exhaustive, once the prior probability for E_1 is established the prior probability for E_2 must logically be given by

$$(2)\quad P(E_2) = 1 - P(E_1)$$

Similarly there are only two likelihoods. $P(E|E_1)$ represents the probability that a student who will be successful would obtain a specific score (where E denotes the event of obtaining that score) on the predictive measure employed. $P(E|E_2)$ represents the probability that a student who will not be successful would obtain that same score on the predictive measure.

Again, logically,

(3) $P(E|E_2) = 1 - P(E|E_1)$

If Bayes' formula, in this context, is written out, we have

$$(4)\quad P(E_1|E) = \frac{P(E_1)P(E|E_1)}{P(E_1)P(E|E_1) + P(E_2)P(E|E_2)}$$

or, making the substitutions indicated by 2 and 3

$$(5)\quad P(E_1|E) = \frac{P(E_1)P(E|E_1)}{P(E_1)P(E|E_1) + [1 - P(E_1)][1 - P(E|E_1)]}$$

and simply letting $P(S) = P(E_1|E)$, $x = P(E_1)$ and $y = P(E|E_1)$ we arrive at

$$(6)\quad P(S) = \frac{xy}{xy + (1 - x)(1 - y)}$$

The posterior probability of success, P(S), is probably all that will be desired for this usage of the formula. However, if the posterior probability of failure is desired it is readily calculated as 1 - P(S).

Bibliography

Lindley, D. V. "A Bayesian solution for some educational prediction problems." Research Bulletin 69-57. Princeton, New Jersey: Educational Testing Service, 1969a.

Lindley, D. V. "A Bayesian solution for some educational prediction problems, II." Research Bulletin 69-91. Princeton, New Jersey: Educational Testing Service, 1970.

Lindley, D. V. "A Bayesian solution for some educational prediction problems, III." Research Bulletin 70-33. Princeton, New Jersey: Educational Testing Service, 1970.

Novick, M. R. and Jackson, P. H. "Bayesian guidance technology." Research Bulletin 69-83. Princeton, New Jersey: Educational Testing Service, 1969.

Novick, M. R., Jackson, P. H., Thayer, D. F. and Cole, N. S. "Applications of Bayesian methods to the prediction of educational performance." Research Bulletin 71-18. Princeton, New Jersey: Educational Testing Service, 1971.

Owen, R. J. "A Bayesian analysis of Rusch's multiplicative Poisson model for misreadings." Research Bulletin 69-64. Princeton: Educational Testing Service, 1969.